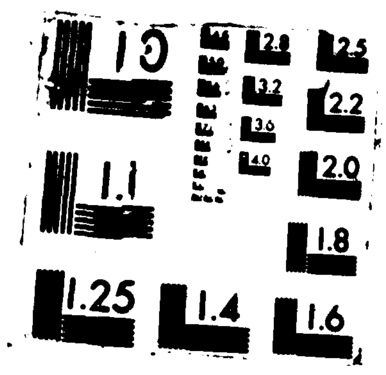
AD-A183 792 AN O(N(3)L) PRIMAL-DUAL INTERIOR POINT ALGORITHM FOR LINEAR PROGRAMMING(U) CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER R C MONTEIRO ET AL. MAY 87 1/1

UNCLASSIFIED ORC-87-4 N00014-87-K-0202 F/G 12/4 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(12)





AN $O(n^3L)$ PRIMAL-DUAL INTERIOR POINT ALGORITHM FOR LINEAR PROGRAMMING*

by

R. C. Monteiro**
I. Adler**

ORC 87-4 May 1987



* This research was sponsored in part by the United States Navy Office of Naval Research under contract N00014-87-K-0202.

** Department of Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, California, 94720.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | | |
|---|---|---|
| 1. REPORT NUMBER<br>ORC 87-4 | 2. GOVT ACCESSION NO.<br>AD-A1[illegible] | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>An $O(n^3L)$ Primal-Dual Interior Point Algorithm for Linear Programming | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>87-4 |
| 7. AUTHOR(s)<br>R. C. Monteiro<br>I. Adler | | 8. CONTRACT OR GRANT NUMBER(s)<br>N000 14-87-K-0202 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Operations Research Center<br>3115 Etcheverry Hall<br>U. C. Berkeley Berkeley, CA 94720 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Dept. of the Navy<br>Arlington, VA 22217 | | 12. REPORT DATE<br>May 1987 |
| | | 13. NUMBER OF PAGES<br>32 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Distribution Unlimited; Approved for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see report)

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

-1-

# An $O(n^3L)$ Primal-Dual Interior Point Algorithm for Linear Programming

R. C. Monteiro and I. Adler

Department of Industrial Engineering and Operations Research
University of California
Berkeley, CA 94720

May 1987

cubed

The authors                    square root of nL.

**Abstract** – ~~We~~ describe a primal-dual interior point algorithm for linear programming problems which requires a total of $O(n^3L)$ arithmetic operations. Each iteration updates a penalty parameter and finds an approximate Newton's direction associated with the Kuhn-Tucker system of equations which characterizes a solution of the logarithm barrier function problem. This direction is then used to find the next iterate. The algorithm is based on the path following idea. The total number of iterations is shown to be of the order of $O(\sqrt{nL})$. Keywords: Convergence

**Key Words** – Interior-point methods, Linear Programming, Karmarkar's LP algorithm, Polynomial-time algorithms, Barrier function, Path following,



## 1. Introduction

Consider the linear programming problem

$$(P) \quad \min \; c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geq 0$$

where $A \in \mathbb{R}^{m \times n}$. This paper presents an algorithm for Linear Programming (LP) problems based on the logarithmic barrier function approach. The logarithmic barrier function method was first used for LP problems by Frisch [1]. The introduction of the new interior point algorithm by Karmarkar [4] led researchers to reconsider the application of the logarithmic barrier function method to LP problems. Recently, this method was first considered by Gill et al. [2] to develop a projected Newton barrier method for solving LP problems.

Meggido [6] provides a theoretical analysis for the logarithmic barrier method and proposes a primal-dual framework based on a consideration of a pair of primal and dual LP problems. Kojima et al. [5], using this framework, present an algorithm that works simultaneously with a pair of primal and dual LP problems. Their algorithm is shown to converge in at most $O(nL)$ iterations with a computational effort of $O(n^3)$ arithmetic operations per iteration, resulting in a total of $O(n^4L)$ arithmetic operations.

In this paper, we build on the ideas in [5] and obtain a faster algorithm. The primal-dual framework presented in [6] is used. The directions generated by our algorithm are essentially the same as the directions generated by the algorithm of Kojima et al. [5]. However, working closer to the "path of solutions" (c.f. [5]), we are able to obtain convergence in at most $O(\sqrt{n}L)$ iterations. Each iteration involves the inversion of a $n \times n$ matrix which can be done in at most $O(n^3)$ arithmetic operations. Based on ideas presented by Gonzaga [3], we are able to exploit the special structure of the matrix to be inverted so that it can be done in an average of $O(n^{2.5})$ arithmetic operations per iteration. Thus overall our algorithm requires $O(n^3L)$ arithmetic operations. It should be noted that Renegar [7] was the first to introduce an interior point algorithm requiring $O(\sqrt{n}L)$ iterations and $O(n^{3.5}L)$ arithmetic operations. Subsequently, Vaidya [8] improved it so that the total complexity is $O(n^3L)$ arithmetic operations. Equivalent complexity was also obtained by an algorithm which was presented by Gonzaga [3]. Both Vaidya's and Gonzaga's algorithms are primal algorithms. It should be noted that in order to simplify the complexity analysis presentation, we assume throughout the paper that $m = O(n)$.

Our paper is organized as follows. In section 2, we present some theoretical background. In section 3, we present the algorithm. In section 4, we prove results related to the convergence properties of the algorithm. In section 5, we present the updating scheme that leads to a reduction in the average number of arithmetic operations per step. In section 6, we discuss how to initialize the algorithm. Finally, we discuss in section 7 some

extensions, and in particular how to extend our algorithm to solve convex quadratic programming in $O(\sqrt{n}L)$ iterations.

## 2. Theoretical Background

In order to facilitate the reading of this paper, we use a notation roughly similar to the one in [5]. A discussion of the main results necessary for the development of our algorithm is presented in this section. These results are adapted from [5]. A detailed discussion of these results can be found in [6].

We consider the pair of the standard form linear program and its dual

$$(P) \quad \min \ c^T x$$
$$s.t. \ Ax = b$$
$$x \geq 0$$

$$(D) \quad \max \ b^T y$$
$$s.t. \ A^T y + z = c$$
$$z \geq 0$$

where $A \in \mathbf{R}^{m \times n}$, $c \in \mathbf{R}^n$ and $b \in \mathbf{R}^m$. We impose the following assumptions:

**Assumptions 2.1:**

(a) The set $S \equiv \left\{ x \in \mathbf{R}^n \, ; \, Ax = b, \, x > 0 \right\}$ is non-empty.

(b) The set $T \equiv \left\{ (y, z) \in \mathbf{R}^{m \times n} \, ; \, A^T y + z = c, \, z > 0 \right\}$ is non-empty.

(c) rank (A) = m.

Throughout this paper, we will denote a point in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ by the lower case letter w, i.e.,

$$w \equiv (x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$$

The logarithmic barrier function technique, usually employed in non-linear constrained optimization, will be applied to the problem (P). The method consists of a consideration of the family of problems

$$(P\mu) \quad \min \; c^T x - \mu \sum_{j=1}^{n} \ln x_j$$

$$s.t. \; Ax = b$$

$$x > 0$$

where $\mu > 0$ is the penalty parameter. As $\mu \to 0$, we would expect the optimal solutions of $(P\mu)$ to converge to an optimal solution of (P). This method, usually attributed to Frisch [1], recently came up in [2] where an analogy with Karmarkar's algorithm is presented. In [6], a comprehensive analysis of this approach is presented, where the problems (P) and (D) interplay together.

From now on, a capital letter corresponding to a lower case letter which denotes a vector, say $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, will denote the diagonal matrix with the components of the vector on the diagonal, i.e., $X = diag(x_1, \ldots, x_n)$. Observe that the objective function of the problem $(P\mu)$ is a strictly convex function. This implies that the problem $(P\mu)$ has at most one global minimum, and that this global minimum, if it exists, is completely characterized by the Kuhn-Tucker stationary condition:

$$(i) \;\; ZXe - \mu e = 0$$

$$(ii) \;\; Ax - b = 0, \; x > 0 \qquad (2.1)$$

$$(iii) \;\; A^T y + z - c = 0$$

where $e \in \mathbb{R}^n$ denotes the vector of ones. In fact, we have the following result:

**Proposition 2.2 :** *Under the assumptions (a) and (b), problem ($P\mu$) (and consequently the system (2.1)) has a unique global solution $x(\mu)$, for all $\mu > 0$.*

Observe that in the system (2.1), $x$ uniquely determines $z$ from the first equation and $y$ from the third equation, since the matrix A has full rank. For each $\mu > 0$, we denote the triple of solutions of (2.1) by $w(\mu) \equiv (x(\mu), y(\mu), z(\mu))$. Obviously $w(\mu) \in S \times T$. Given a point $w = (x, y, z) \in S \times T$, we define the duality gap at w to be

$$g(w) = c^T x - b^T y$$

Using the two last equations in (2.1), we can easily verify that

$$g(w) = x^T z \quad , \quad w \in S \times T \tag{2.2}$$

In particular, using the first equation in (2.1), we obtain $g(w(\mu)) = n\mu$, for all $\mu$ and therefore $g(w(\mu))$ converges to zero as $\mu$ goes to zero. This implies that $c^T x(\mu)$ and $b^T y(\mu)$ converges to the optimal value of the problems (P) and (D) respectively. In fact, we have the following stronger result:

**Proposition 2.3 :** *Under assumption 2.1, as $\mu \rightarrow 0$, $x(\mu)$ ( $(y(\mu), z(\mu))$ ) converges to an optimal solution of problem (P) ((D)).*

The following notation will be useful later. Let $w \in S \times T$. We denote by $f(w) \equiv (f_1(w), \ldots, f_n(w))^T \in \mathbb{R}^n$ the n-vector defined by

$$f_i \equiv x_i z_i \quad , \quad i = 1, \ldots, n$$

With this notation, the first equation of (2.1) written coordinate-wise becomes :

$$f_i(w(\mu)) \equiv x_i(\mu)\, z_i(\mu) = \mu \;, \quad i = 1,\ldots,n \tag{2.3}$$

We denote by $\Gamma$ the path of solutions $w(\mu)$ , $\mu > 0$, i.e.,

$$\Gamma \equiv \left\{ w(\mu) \equiv (x(\mu), y(\mu), z(\mu)) \;;\; \mu > 0 \right\}.$$

The algorithm which will be presented in the next section, will "follow" this path $\Gamma$ with the objective of approaching the desired solutions of the original problems (P) and (D). The path following procedure is described in the next sections.

## 3. The Algorithm

As in the previous section, we denote a triple $(x, y, z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ by the lower case letter w, i.e., $w \equiv (x, y, z)$. The algorithm will generate a sequence of points $w^k \in S \times T$, k = 0,1,2,.... where the initial point $w^o$ is provided as input to the algorithm. In this section, we require that $w^o \in S \times T$ be a point satisfying some criterion of closeness with respect to the path of solutions $\Gamma$. Given an LP problem in standard form, in section 6 ' e show how to construct an equivalent LP problem so that assumption 2.1 is satisfied. As a consequence of this construction, we also show how to obtain an initial point $w^o \in S \times T$ satisfying the criterion of closeness.

Given a current iterate $(x, y, z) \in S \times T$, a triple of directions $(\Delta x, \Delta y, \Delta z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ needs to be generated for the determination of the next iterate. Throughout this paper, a triple of directions $(\Delta x, \Delta y, \Delta z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ is denoted by the symbol $\Delta w$. Let $(\hat{x}, \hat{y}, \hat{z})$ denote the next iterate. We obtain $(\hat{x}, \hat{y}, \hat{z})$ by

$$\hat{x} \equiv x - \Delta x$$

$$\hat{y} \equiv y - \Delta y$$

$$\hat{z} \equiv z - \Delta z$$

or in more compact notation

$$\hat{w} \equiv w - \Delta w$$

According to [5], the direction $\Delta w$ chosen to generate the next iterate $\hat{w}$, is defined as the Newton's direction associated with the Kuhn-Tucker system of equations (2.1).

However, with the objective of improving the worst-case complexity on the number of arithmetic operations, we consider a slight variation of the direction used in [5]. If we denote the left hand side of the system of equations (2.1) by $H(w) \equiv H(x, y, z)$, the Newton's direction $\Delta w$ at $w \in S \times T$ is defined by the system of linear equations

$$D_w H(w)\, \Delta w = H(w)$$

where $\Delta w = (\Delta x, \Delta y, \Delta z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ and $D_w H(w)$ denotes the Jacobian of $H$ at $w \equiv (x, y, z)$. We observe that $D_w H(x, y, z)$ does not depend on the argument $y \in \mathbf{R}^m$. Indeed, the Jacobian of $H$ at $w \equiv (x, y, z)$ is given by

$$J(x, z) \equiv D_w H(w) = \begin{bmatrix} Z & 0 & X \\ A & 0 & 0 \\ 0 & A^T & I \end{bmatrix}$$

The direction $\Delta w$ that we are going to consider is defined by the following system of linear equations

$$J(\tilde{x}, \tilde{z})\, \Delta w = H(x, y, z)$$

where the points $\tilde{x} \in \mathbf{R}^n$ and $\tilde{z} \in \mathbf{R}^n$ will be chosen to approximate $x \in \mathbf{R}^n$ and $z \in \mathbf{R}^n$ respectively in a manner which will be specified latter. More specifically, $\Delta w = (\Delta x, \Delta y, \Delta z)$ is defined by the following system of linear equations

$$\tilde{Z}\Delta x + \tilde{X}\Delta z = XZe - \hat{\mu} e \tag{3.1.a}$$

$$A\Delta x = 0 \tag{3.1.b}$$

$$A^T\Delta y + \Delta z = 0 \tag{3.1.c}$$

where $\hat{\mu} > 0$ is some prespecified penalty parameter. Observe that the right hand sides of (3.1.b) and (3.1.c) are zero since w is assumed to be in the set $S \times T$. However the right hand side of (3.1.a) is not necessarily zero and equals zero only when the point w lies on the path $\Gamma$. Throughout this paper, the lower case letter s will denote a pair

$(x, z) \in \mathbf{R}^n \times \mathbf{R}^n$. Note that the solution $\Delta w = (\Delta x, \Delta y, \Delta z)$ of the system of equations (3.1) clearly depends on the current iterate $w = (x, y, z)$, on the Jacobian of $H$ at the "approximation" $\tilde{s} = (\tilde{x}, \tilde{z})$ of $s = (x, z)$ , and on the penalty parameter $\hat{\mu} > 0$. In order to indicate this dependence, we denote the solution $(\Delta x, \Delta y, \Delta z)$ of the system (3.1) by

$$\Delta w(w, \tilde{s}, \hat{\mu})$$

By simple calculation, we obtain the following expressions for $\Delta x$, $\Delta y$, $\Delta z$

$$\Delta x = \left[\tilde{Z}^{-1} - \tilde{Z}^{-1}\tilde{X}A^T(A\tilde{Z}^{-1}\tilde{X}A^T)^{-1}A\tilde{Z}^{-1}\right](XZe - \hat{\mu}e)$$

$$\Delta y = -\left[(A\tilde{Z}^{-1}\tilde{X}A^T)^{-1}A\tilde{Z}^{-1}\right](XZe - \hat{\mu}e)$$

$$\Delta z = \left[A^T (A\tilde{Z}^{-1}\tilde{X}A^T)^{-1}A\tilde{Z}^{-1}\right](XZe - \hat{\mu}e)$$

Therefore, to calculate the direction $\Delta w \equiv (\Delta x, \Delta y, \Delta z)$, the inverse of the matrix $(A\tilde{Z}^{-1}\tilde{X}A^T)$ needs to be calculated. This is the main motivation to consider just an approximation $\tilde{s} = (\tilde{x}, \tilde{z})$ of $s = (x, z)$ so that we do not need to invert this matrix from scratch at every iteration. If the current diagonal matrix $\tilde{Z}^{-1}\tilde{X}$ differs from the previous one by exactly $l$ diagonal elements then, by performing $l$ rank-one updates, we are able to compute the inverse of the matrix $(A\tilde{Z}^{-1}\tilde{X}A^T)$ in $O(n^2 l)$ arithmetic operations. Observe that all the other operations involved in the computation of $\Delta w \equiv \Delta w(w, \tilde{s}, \hat{\mu})$ is of the order of $O(n^2)$ arithmetic operations.

We are now ready to describe the algorithm. At the beginning of the algorithm, we assume that an initial point $w^o \equiv (x^o, y^o, z^o) \in S \times T$ is available such that the following criterion of closeness with respect to the path $\Gamma$ is satisfied:

$$\| f(w^o) - \mu^o e \| \le \theta\mu^o \tag{3.2}$$

where $\|\ .\ \|$ denotes the Euclidean norm, $\mu^o$ is a positive constant and $\theta = 0.1$.

We now state the algorithm.

**Algorithm 3.1 :**

**Step 0)** *Let $w^o \in S \times T$ and $\mu^o > 0$ satisfy (3.2). Let $\epsilon$ be a tolerance for the duality gap. Let*

$$\begin{aligned} \theta &:= 0.1 \\ \delta &:= 0.1 \\ \gamma &:= 0.1 \end{aligned} \tag{3.3}$$

*Set $k:=0$.*

**Step 1)** *If $c^T x^k - b^T y^k < \epsilon$, stop.*

**Step 2)** *Choose $\tilde{s} = (\tilde{x}, \tilde{z})$ in $\mathbf{R}^n_+ \times \mathbf{R}^n_+$. satisfying:*

$$\frac{|\ x_i^k - \tilde{x}_i\ |}{|\ \tilde{x}_i\ |} \le \gamma\ ,\quad i = 1, \ldots, n$$

$$\frac{|\ z_i^k - \tilde{z}_i\ |}{|\ \tilde{z}_i\ |} \le \gamma\ ,\quad i = 1, \ldots, n$$

**Step 3)** *Set $\mu^{k+1} := \mu^k(1 - \delta/\sqrt{n})$.*

*Calculate $\Delta w^k \equiv \Delta w(w^k, \tilde{s}, \mu^{k+1})$.*

**Step 4)** *Set $w^{k+1} := w^k - \Delta w^k$.*

*Set $k := k + 1$ and go to step 1.*

In the following sections, we prove that the algorithm above is a valid one in the sense that it generates at every iteration a point $w^k$ in the set $S \times T$. We also show that it terminates in at most $O(\sqrt{n}\ \max(\log\epsilon^{-1}, \log n, \log\mu^o))$ iterations. Finally, we present a

suitable choice for the approximation point $\bar{s} = (\bar{x}, \bar{z})$ (see step 2 of the algorithm) that will enable us to show that algorithm 3.1 solves the pair of problems (P) and (D) in no more than $O(n^3 \max(\log\epsilon^{-1}, \log n, \log\mu^o))$ arithmetic operations.

## 4. Convergence Results

We begin this section by stating the main result and its consequences. Given two vectors $x \in \mathbf{R}^n$ and $\bar{x} \in \mathbf{R}^n$, we denote the Euclidean norm of the vector $X^{-1}(\bar{x} - x)$ by $\|\bar{x} - x\|_x$, i.e.,

$$\| x - x\|_x = \left[ \sum_{i=1}^{n} \left( \frac{\bar{x}_i - x_i}{x_i} \right)^2 \right]^{1/2} \tag{4.1}$$

The main result is:

**Theorem 4.1**: *Let* $w = (x, y, z) \in S \times T$ *and* $\mu > 0$ *satisfy*

$$\| f(w) - \mu e \| \leq \theta\mu \tag{4.2}$$

*Let* $\bar{s} = (\bar{x}, \bar{z}) \in \mathbf{R}^n_+ \times \mathbf{R}^n_+$ *satisfy*

$$\frac{| x_i - \bar{x}_i |}{| \bar{x}_i |} \leq \gamma \ , \ i = 1, \ldots, n \tag{4.3}$$

$$\frac{| z_i - \bar{z}_i |}{| \bar{z}_i |} \leq \gamma \ , \ i = 1, \ldots, n \tag{4.4}$$

*Let* $\hat{\mu} > 0$ *be defined as*

$$\hat{\mu} = \mu(1 - \delta /\sqrt{n}) \tag{4.5}$$

*Consider the point* $\hat{w} \equiv (\hat{x}, \hat{y}, \hat{z}) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ *defined by*

$$\hat{w} \equiv w - \Delta w \tag{4.6}$$

*where* $\Delta w \equiv \Delta w(w, \tilde{s}, \hat{\mu})$. *Then the following holds:*

*(a) The point* $\hat{w}$ *is in* $S \times T$ *and satisfies*

$$\|\hat{x} - x\|_x \le 0.28 \tag{4.7}$$

$$\|\hat{z} - z\|_z \le 0.28 \tag{4.8}$$

(b) $\|f(\hat{w}) - \hat{\mu}e\| \le \theta\hat{\mu}$

(c) $g(\hat{w}) \equiv c^T\hat{x} - b^T\hat{y} \le 1.1n\hat{\mu}$

The proof of theorem 4.1 will be given at the end of this section. Suppose $w(\mu) \equiv (x(\mu), y(\mu), z(\mu))$ is the point in the path $\Gamma$ corresponding to the penalty parameter $\mu$. The criterion of closeness (4.2) is then equivalent to

$$\| f(w) - f(w(\mu)) \| \le \theta\mu \tag{4.9}$$

since $f(w(\mu)) = \mu e$ by relation (2.3). By theorem 4.1, relation (4.9) will hold for the new point $\hat{w}$ defined by (4.6) and the penalty parameter $\hat{\mu} > 0$ given by (4.5).

As a consequence of theorem 4.1, we have the following result:

**Corollary 4.2:** *All points* $w^k$ *generated by algorithm 3.1 satisfy*

(a) $w^k$ is in $S \times T$, for all $k = 1,2, \dots$ and

$$\| x^{k+1} - x^k \|_{x^k} \le 0.28$$

$$\| z^{k+1} - z^k \|_{z^k} \le 0.28$$

(b) $\| f(w^k) - \mu^k e \| \le \theta\mu^k$ , for all $k = 1,2, \dots$

(c) $g(w^k) \equiv c^T x^k - b^T y^k \le 1.1n\mu^k$ , for all $k = 1,2, \dots$

*where*

$$\mu^k = \mu^o( 1 - \delta /\sqrt{n} )^k \quad \text{for} \quad k = 1,2, \dots$$

**Proof:** This result follows trivially by arguing inductively and using theorem 4.1. □

By (b) of corollary 4.2, all iterates $w^k$ generated by algorithm 3.1 will satisfy the criterion of closeness

$$\| f(w^k) - f(w(\mu^k)) \| \leq \theta\mu^k$$

Therefore we can view algorithm 3.1 as a path-following procedure, i.e., the iterates $w^k$ tries to trace the path $\Gamma$ so that it eventually converges to an optimal solution $w^* = (x^*, y^*, z^*)$ for the pair of LP problems (P) and (D).

We now derive an upper bound on the total number of iterations performed by algorithm 3.1.

**Proposition 4.3:** *The total number of iterations performed by algorithm 3.1 is no greater than* $k^* \equiv \lceil \log(1.1n\epsilon^{-1}\mu^o)\sqrt{n} / \delta \rceil$ *where* $\epsilon > 0$ *denotes the tolerance for the duality gap and* $\mu^o$ *is the initial penalty parameter.*

**Proof:** From (c) of corollary 4.2, we can terminate the algorithm as soon as

$$1.1n\mu^k \leq \epsilon$$

It is enough to show that $k^*$ satisfies the inequality above. By the definition of $k^*$, we have

$$\log \epsilon \geq -\frac{k^*\delta}{\sqrt{n}} + \log(1.1n\mu^o)$$

$$\geq k^* \log\left(1 - \frac{\delta}{\sqrt{n}}\right) + \log(1.1n\mu^o)$$

$$= \log\left[1.1n\mu^o\left(1 - \frac{\delta}{\sqrt{n}}\right)^{k^*}\right]$$

$$= \log 1.1n\mu^{k^*}$$

Second inequality is due to the fact that $\log(1 - x) \leq -x$ for all $x > -1$ and the last equality follows from the definition of $\mu^{k^*}$. Therefore $k^*$ satisfies

$$1.1 n\mu^{k^*} \le \varepsilon$$

and this completes the proof of the proposition. □

Let $L$ denote the number of bits to represent the data of the primal LP problem (P), i.e., L is the size of the problem (P). It is a well-known fact that if, at some iteration $k$, the duality gap satisfies $c^T x^k - b^T y^k \le 2^{-L}$, then we may obtain optimal solutions for problems (P) and (D) in $O(mn^2)$ arithmetic operations.

Using this observation, we obtain

**Corollary 4.4:** *If the initial penalty parameter $\mu^o$ satisfies* $\log \mu^o = O(L)$ *then algorithm 3.1 solves the LP problem (P) in at most $O(\sqrt{n}L)$ iterations.*

**Proof:** Follows directly from the previous proposition. □

In section 6, we will see that the initial penalty parameter $\mu^o$ can be chosen to satisfy $\log \mu^o = O(L)$. One possible choice for the approximation $\tilde{s} \equiv (\tilde{x}, \tilde{z})$ on step 2 of the algorithm is to use exact data, that is, to set $\tilde{s}$, on the $k^{th}$ iteration, equal to $s^k$. With this choice of the "approximation" $\tilde{s}$ , we have the following result:

**Corollary 4.5:** *Algorithm 3.1 solves the pair of LP problems (P) and (D) in no more than $O(n^{3.5}L)$ iterations.*

**Proof:** At every iteration, we need to calculate the inverse of the matrix $[A(Z^k)^{-1}X^kA^T]$ and this requires $O(n^3)$ arithmetic operations. By corollary 4.4, algorithm 3.1 terminates in at most $O(\sqrt{n}L)$ iterations. These two observations immediately concludes the proof of the corollary. □

In the next section, we present an updating scheme for the approximation $\tilde{s}$ that will reduce the worst-case complexity of algorithm 3.1 to $O(n^3L)$ arithmetic operations.

We now concentrate our effort towards proving theorem 4.1. We should point out that some of the arguments below become simpler when we use exact data for the

approximation $\tilde{s}$.

Let $w = (x, y, z) \in S \times T$, $\tilde{s} = (\tilde{x}, \tilde{z}) \in \mathbb{R}^n \times \mathbb{R}^n$ and $\hat{\mu} > 0$. Let $\Delta w = (\Delta x, \Delta y, \Delta z)$ be the direction $\Delta w(w, \tilde{s}, \hat{\mu})$. Consider the point $\hat{w}$ defined as in relation (4.6). The next result provides expressions for the product of complementary variables $f_i(\hat{w})$ , $i = 1, \ldots, n$ and the duality gap $g(\hat{w}) \equiv c^T\hat{x} - b^T\hat{y}$.

**Proposition 4.6:** *Let $w$ ,$\tilde{s}$ and $\hat{w}$ be as above. Then the following expressions hold:*

$$f_i(\hat{w}) = \hat{\mu} + \Delta x_i \Delta z_i + (\tilde{x}_i - x_i)\Delta z_i + (\tilde{z}_i - z_i)\Delta x_i \tag{4.10}$$

$$(\Delta x)^T(\Delta z) = 0 \tag{4.11}$$

$$g(\hat{w}) = \sum_{i=1}^{n} f_i(\hat{w}) = n\hat{\mu} + (\tilde{z} - z)^T\Delta x + (\tilde{x} - x)^T\Delta z \tag{4.12}$$

**Proof:** By definition, we have for all i = 1,...,n

$$\begin{aligned} f_i(\hat{w}) &\equiv \hat{x}_i \hat{z}_i \\ &= (x_i - \Delta x_i)(z_i - \Delta z_i) \\ &= x_i z_i - (x_i \Delta z_i + z_i \Delta x_i) + \Delta x_i \Delta z_i \\ &= x_i z_i - (\tilde{x}_i \Delta z_i + \tilde{z}_i \Delta x_i) + \Delta x_i \Delta z_i + (\tilde{x}_i - x_i)\Delta z_i + (\tilde{z}_i - z_i)\Delta x_i \end{aligned} \tag{4.13}$$

Writing (3.1.a) coordinate-wise, we obtain for all i = 1,...,n

$$(\tilde{x}_i \Delta z_i + \tilde{z}_i \Delta x_i) = x_i z_i - \hat{\mu} \tag{4.14}$$

Relations (4.13) and (4.14) immediately imply (4.10). Now we prove (4.11). Multiplying expression (3.1.c) on the left by $(\Delta x)^T$, we obtain

$$(A\Delta x)^T \Delta y + (\Delta x)^T \Delta z = 0 \tag{4.15}$$

Relations (3.1.b) and (4.15) immediately imply (4.11). Note that the first equality in (4.12) follows immediately from expression (2.2) and the definition of $f_i(\hat{w})$. Summing expres-

sion (4.10) over all indices i = 1,...,n and noting (4.11), we immediately obtain the second equality in (4.12). This completes the proof of the proposition. □

Observe that if $\tilde{x} = x$ and $\tilde{z} = z$, that is, the approximation is exact, then the duality gap at $\hat{w}$ becomes $g(\hat{w}) = n\hat{\mu}$. In this case, (c) of theorem 4.1 follows trivially.

We now state and prove some preliminary results that will be useful in the proof of Theorem 4.1.

Let $w = (x, y, z) \in S \times T$, $\tilde{s} = (\tilde{x}, \tilde{z}) \in \mathbf{R}^n \times \mathbf{R}^n$ and $\hat{\mu} > 0$. Let $\Delta w = (\Delta x, \Delta y, \Delta z)$ be the direction $\Delta w(w, \tilde{s}, \hat{\mu})$. We denote by $\Delta f \equiv (\Delta f_1, \ldots, \Delta f_n)^T$ the n-vector defined as

$$\Delta f = (\Delta x_1 \Delta z_1, \ldots, \Delta x_n \Delta z_n)^T \tag{4.16}$$

where $\Delta x_i$ and $\Delta z_i$ denotes the $i^{th}$ coordinate of the vectors $\Delta x$ and $\Delta z$ respectively. The next results provides an upper bound on the Euclidean norm of the vector $\Delta f$.

**Lemma 4.7**: *Let $\Delta f$ be defined as in (4.16). Then, we have*

$$\| \Delta f \| \le \frac{\|f(w) - \hat{\mu} e\|^2}{2\tilde{f}_{\min}} \tag{4.17}$$

where

$$\tilde{f}_{\min} \equiv \min \left\{ \mid \tilde{x}_i \tilde{z}_i \mid \; ; \; i = 1, \ldots, n \right\} \tag{4.18}$$

*Furthermore, we have*

$$\|\tilde{D} \Delta z\|^2 \le \frac{\|f(w) - \hat{\mu} e\|^2}{\tilde{f}_{\min}} \tag{4.19}$$

$$\|\tilde{D}^{-1} \Delta x\|^2 \le \frac{\|f(w) - \hat{\mu} e\|^2}{\tilde{f}_{\min}} \tag{4.20}$$

***where $\tilde{D}$ is the diagonal matrix defined by***

$$\tilde{D} = (\tilde{Z}^{-1}\tilde{X})^{1/2} \tag{4.21}$$

**Proof:** By equation (3.1.a), we have

$$\tilde{D}^{-1}\Delta x + \tilde{D}\Delta z = (\tilde{X}\tilde{Z})^{1/2}(XZ - \hat{\mu}e) \tag{4.22}$$

From (4.11) it follows that

$$(\tilde{D}^{-1}\Delta x)^T(\tilde{D}\Delta z) = 0 \tag{4.23}$$

Using the Pythagorean theorem, relations (4.22), (4.23) and the definition of the Euclidean norm, we obtain

$$\begin{aligned} \| \tilde{D}^{-1}\Delta x \|^2 + \| \tilde{D}\Delta z \|^2 &= \| (\tilde{X}\tilde{Z})^{1/2}(XZe - \hat{\mu}e) \|^2 \\ &= \sum_{i=1}^{n} \frac{(f_i(w) - \hat{\mu})^2}{|\tilde{x}_i\tilde{z}_i|} \\ &\le \frac{\|f(w) - \hat{\mu}e\|^2}{\tilde{f}_{\min}} \end{aligned} \tag{4.24}$$

Inequalities (4.19) and (4.20) follow immediately from (4.24). Also (4.24) implies that

$$\| \tilde{D}^{-1}\Delta x \| \, \| \tilde{D}\Delta z \| \le \frac{\|f(w) - \hat{\mu}e\|^2}{2\tilde{f}_{\min}} \tag{4.25}$$

On the other hand, using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \sum_{i=1}^{n} | \Delta x_i \Delta z_i | &= \sum_{i=1}^{n} | \tilde{D}_{ii}^{-1}\Delta x_i | \, | \tilde{D}_{ii}\Delta z_i | \\ &\le \| \tilde{D}^{-1}\Delta x \| \, \| \tilde{D}\Delta z \| \end{aligned} \tag{4.26}$$

Since $\| \Delta f \| \le \sum_{i=1}^{n} | \Delta x_i \Delta z_i |$, relations (4.25) and (4.26) imply the inequality (4.17). This completes the proof of the lemma. □

The next lemma provides important relations that will be useful in the proof of theorem 4.1. We observe that the specialization of the results below to the case when the approximation becomes exact simply involves substitution of $\tilde{x}$ by $x$, $\tilde{z}$ by $z$ and $\gamma$ by 0.

**Lemma 4.8:** *Let* $w = (x, y, z) \in S \times T$, $\tilde{s} = (\tilde{x}, \tilde{z}) \in \mathbf{R}^n_+ \times \mathbf{R}^n_+$, $\mu > 0$ *and* $\hat{\mu} > 0$ *be as in the statement of theorem 4.1. Let* $\Delta w \equiv (\Delta x, \Delta y, \Delta z)$ *be the direction* $\Delta w(w, \tilde{s}, \hat{\mu})$. *Let* $p$ *and* $q$ *be defined by*

$$p = (1 - \theta)(1 + \gamma)^{-2} \tag{4.27}$$

$$q = (1 + \theta)(1 - \gamma)^{-2} \tag{4.28}$$

*Then the following relations hold:*

$$p\mu \le \tilde{x}_i \tilde{z}_i \le q\mu \quad , i = 1,\dots,n \tag{4.29}$$

$$\| \Delta f \| \le \frac{(\theta + \delta)^2 \mu}{2p} \tag{4.30}$$

$$\| \tilde{D}^{-1} \Delta x \|^2 \le \frac{(\theta + \delta)^2 \mu}{p} \tag{4.31}$$

$$\| \tilde{D} \Delta z \|^2 \le \frac{(\theta + \delta)^2 \mu}{p} \tag{4.32}$$

*where* $\tilde{D}$ *is defined by (4.21).*

**Proof:** From (4.3) and (4.4) it follows that

$$0 < 1 - \gamma \le \frac{x_i}{\tilde{x}_i} \le 1 + \gamma \ , \ i = 1,\dots,n \tag{4.33}$$

$$0 < 1 - \gamma \le \frac{z_i}{\tilde{z}_i} \le 1 + \gamma \ , \ i = 1,\dots,n \tag{4.34}$$

which implies,

$$(1 - \gamma)^2 \le \frac{x_i z_i}{\tilde{x}_i \tilde{z}_i} \le (1 + \gamma)^2 \ , i = 1,\dots,n$$

or equivalently, for all i=1,...,n

$$(1+\gamma)^{-2}x_i z_i \le \tilde{x}_i \tilde{z}_i \le (1-\gamma)^{-2}x_i z_i \tag{4.35}$$

Using (4.2), we obtain

$$(1-\theta)\mu \le x_i z_i \le (1+\theta)\mu \quad , i = 1, \ldots, n \tag{4.36}$$

Relations (4.35) and (4.36) imply that for all i=1,...,n

$$(1+\gamma)^{-2}(1-\theta)\mu \le \tilde{x}_i \tilde{z}_i \le (1-\gamma)^{-2}(1+\theta)\mu$$

which is exactly (4.29).

Since $\|e\| = \sqrt{n}$, relations (4.2) and (4.5) imply that

$$\begin{aligned} \|f(w) - \hat{\mu} e\|^2 &\le \Big( \| f(w) - \mu e \| + \| \mu e - \hat{\mu} e \| \Big)^2 \\ &\le \Big( \theta\mu + |\mu - \hat{\mu}| \, \|e\| \Big)^2 \\ &\le (\theta\mu + \delta\mu)^2 \\ &\le (\theta + \delta)^2 \mu^2 \end{aligned} \tag{4.37}$$

Using lemma 4.7, relations (4.29) and (4.37) , we immediately obtain (4.30), (4.31) and (4.32). This completes the proof of the lemma. □

We are now ready to prove theorem 4.1.

**Proof of theorem 4.1:**

(a) From (3.1.b), (3.1.c) and the fact that $w \in S \times T$, it follows that $\hat{w} \equiv (\hat{x}, \hat{y}, \hat{z})$ satisfies

$$A \hat{x} = b$$

$$A^T \hat{y} + \hat{z} = c$$

We have just to show that $\hat{x}$ and $\hat{z}$ are strictly positive to conclude that $\hat{w} \in S \times T$. We note that (4.7) and (4.8) immediately imply that $\hat{x} > 0$ and $\hat{z} > 0$ since $x > 0$ and $z > 0$ by

assumption. We will now show that (4.7) and (4.8) hold. By the definition of $\hat{x}$, we obtain

$$\|\hat{x} - x\|_x^2 = \sum_{i=1}^{n} \left( \frac{\Delta x_i}{x_i} \right)^2$$

$$= \sum_{i=1}^{n} \left( \frac{\tilde{D}_{ii}}{x_i} \right)^2 \left( \tilde{D}_{ii}^{-1} \Delta x_i \right)^2 \tag{4.38}$$

where $\tilde{D} = diag(\tilde{D}_{11}, \ldots, \tilde{D}_{nn})$ is given by (4.21), i.e.,

$$\tilde{D}_{ii} = \left( \frac{\tilde{x}_i}{\tilde{z}_i} \right)^{1/2} \quad i = 1, \ldots, n \tag{4.39}$$

Using (4.2), (4.3), (4.4) and the expression for $\tilde{D}_{ii}$ above, we obtain for all $i = 1, \ldots, n$

$$\left( \frac{\tilde{D}_{ii}}{x_i} \right)^2 = \frac{\tilde{x}_i}{\tilde{z}_i x_i^2}$$

$$= \left( \frac{z_i}{\tilde{z}_i} \right) \left( \frac{\tilde{x}_i}{x_i} \right) \left( \frac{1}{x_i z_i} \right)$$

$$\leq \left( \frac{1 + \gamma}{1 - \gamma} \right) \left( \frac{1}{(1 - \theta)\mu} \right) \tag{4.40}$$

The last inequality together with (4.38) implies

$$\|\hat{x} - x\|_x^2 \leq \left( \frac{1 + \gamma}{1 - \gamma} \right) \left[ \frac{1}{(1 - \theta)\mu} \right] \| \tilde{D}^{-1} \Delta x \|^2 \tag{4.41}$$

Inequality (4.31) of lemma 4.8 and (4.41) then imply

$$\|\hat{x} - x\|_x \leq \left[ \frac{(1 + \gamma)}{(1 - \gamma)(1 - \theta)\, p} \right]^{1/2} (\theta + \delta) \tag{4.42}$$

Substituting in (4.42) the values of $\theta$, $\delta$, $\gamma$, and $p$ from (3.3), (4.27) and (4.28) we obtain (4.7). In a similar way, one can prove (4.8). This completes the proof of (a).

(b) Writing expression (4.10) in vector notation and noting (4.16), we obtain

$$f(\hat{w}) - \hat{\mu}e = \Delta f + (\tilde{X} - X)\Delta z + (\tilde{Z} - Z)\Delta x$$

By the properties of norms, we obtain

$$\| f(\hat{w}) - \hat{\mu}e \| \le \| \Delta f \| + \| (\tilde{X} - X)\Delta z \| + \| (\tilde{Z} - Z)\Delta x \| \tag{4.43}$$

Using the definition of the Euclidean norm, we obtain

$$\begin{aligned} \| (\tilde{X} - X)\Delta z \|^2 &= \sum_{i=1}^{n} (\tilde{x}_i - x_i)^2 (\Delta z_i)^2 \\ &= \sum_{i=1}^{n} [\tilde{D}_{ii}^{-1}(\tilde{x}_i - x_i)]^2 (\tilde{D}_{ii}\Delta z_i)^2 \end{aligned} \tag{4.44}$$

where $\tilde{D}_{ii}$ is as in (4.39). Using (4.3), (4.29) and (4.39), we obtain

$$\begin{aligned} [\tilde{D}_{ii}^{-1}(\tilde{x}_i - x_i)]^2 &= (\tilde{x}_i \tilde{z}_i) \left( \frac{\tilde{x}_i - x_i}{\tilde{x}_i} \right)^2 \\ &\le q\mu\gamma^2 \ , \ i = 1, \dots, n \end{aligned} \tag{4.45}$$

In a similar way, one can prove that

$$[\tilde{D}_{ii}(\tilde{z}_i - z_i)]^2 \le q\mu\gamma^2 \ , \ i = 1, \dots, n \tag{4.46}$$

Relation (4.32) of lemma 4.8 and relations (4.44) and (4.45) imply that

$$\| (\tilde{X} - X)\Delta z \|^2 \le \frac{q\gamma^2(\theta + \delta)^2\mu^2}{p} \tag{4.47}$$

In a similar way, one can prove that

$$\| (\tilde{Z} - Z)\Delta x \|^2 \le \frac{q\gamma^2(\theta + \delta)^2\mu^2}{p} \tag{4.48}$$

From (4.43), (4.30), (4.47) and (4.48) it follows that

$$\| f(\hat{w}) - \hat{\mu} e \| \le \left[ \frac{(\theta + \delta)^2}{2p} + 2\gamma(\theta + \delta)(\frac{q}{p})^{\frac{1}{2}} \right] \mu$$

From the expression of $\hat{\mu}$ in (4.5), we obtain

$$\mu \le \frac{\hat{\mu}}{(1 - \delta)}$$

which implies

$$\| f(\hat{w}) - \hat{\mu} e \| \le \frac{1}{(1 - \delta)} \left[ \frac{(\theta + \delta)^2}{2p} + 2\gamma(\theta + \delta)(\frac{q}{p})^{\frac{1}{2}} \right] \hat{\mu}$$

Substituting the values of $\theta$, $\delta$, $\gamma$, $p$, and $q$ from (3.3), (4.27) and (4.28) we obtain

$$\| f(\hat{w}) - \hat{\mu} e \| \le \theta \hat{\mu} \qquad (4.49)$$

and this completes the proof of (b).

(c) The first equality of expression (4.12) says that

$$g(\hat{w}) = e^T f(\hat{w})$$

Using the Cauchy-Schwarz inequality and expression (4.49), we obtain

$$\begin{aligned} g(\hat{w}) &\le \|e\| \, \|f(\hat{w})\| \\ &\le \sqrt{n} \left( \|f(\hat{w}) - \hat{\mu} e\| + \|\hat{\mu} e\| \right) \\ &\le \sqrt{n} \, (\theta \hat{\mu} + \sqrt{n} \hat{\mu}) \\ &\le (1 + \theta) n \hat{\mu} \\ &= 1.1 n \hat{\mu} \end{aligned}$$

since $\theta = 0.1$. This completes the proof of (c). $\square$

## 5. A Good Choice for $\tilde{x}$ and $\tilde{z}$

In this section, we show how to choose the approximation $\tilde{s} \equiv (\tilde{x}, \tilde{z}) \in \mathbf{R}^n \times \mathbf{R}^n$ of $s^k = (x^k, z^k)$ in step 2 of algorithm 3.1 so that its worst-case complexity reduces to the order of $O(n^3 L)$ arithmetic operations.

The choice of the approximation $\tilde{s}$ is made by an updating scheme as follows (In the procedure below, k stands for the iteration count):

**Updating scheme 5.1:**

*For* $k := 0$, *set* $\tilde{x} := x^o$ *and* $\tilde{z} := z^o$

*For* $k > 0$ *do*

*For* $i = 1, \ldots, n$ *do*

*If one of the following holds*

(a) $$\frac{| x_i^k - \tilde{x}_i |}{| \tilde{x}_i |} > \gamma$$

(b) $$\frac{| z_i^k - \tilde{z}_i |}{| \tilde{z}_i |} | > \gamma$$

*then set* $\tilde{x}_i := x_i^k$ *and* $\tilde{z}_i := z_i^k$.

In section 3, we saw that the main effort in each iteration was in computing the inverse of the the matrix $(A\tilde{Z}^{-1}\tilde{X}A^T)$. If only $l$ diagonal elements of the matrix $(\tilde{Z}^{-1}\tilde{X})$ change, this computation can be carried out in $O(n^2 l)$ arithmetic operations by means of $l$ rank-one updates. If we use the updating scheme 5.1 then the $i^{th}$ diagonal element of the matrix $(\tilde{Z}^{-1}\tilde{X})$ changes only when inequality (a) or (b) of scheme 5.1 is satisfied. Next, we provide an upper bound on the total number of diagonal element changes that occurs on the matrix $(\tilde{Z}^{-1}\tilde{X})$ during K iterations of algorithm 3.1. Consider the following two sets:

$$S_i^K = \left\{ k \; ; \; \frac{| x_i^k - \tilde{x}_i |}{| \tilde{x}_i |} > \gamma , \; 1 \le k \le K \right\}$$

$$T_i^K = \left\{ k \; ; \; \frac{| z_i^k - \tilde{z}_i |}{| \tilde{z}_i |} > \gamma , \; 1 \le k \le K \right\}$$

where K is the number of iterations performed. Note that by corollary 4.2, the iterates $w^k = (x^k, y^k, z^k)$ generated by algorithm 3.1 satisfy

$$\|x^{k+1} - x^k\|_{x^k} \le 0.28$$

$$\|z^{k+1} - z^k\|_{z^k} \le 0.28$$

for all integers $k \ge 0$.

The following result can be proved by slightly modifying the arguments in section 5 of [3].

**Proposition 5.2:** *Let $S$ and $T$ be defined as*

$$S = \sum_{i=1}^{n} | S_i^K |$$

$$T = \sum_{i=1}^{n} | T_i^K |$$

*Then, we have*

$$S \le 4.5\sqrt{n}K$$

$$T \le 4.5\sqrt{n}K$$

As a consequence of this result, we have :

**Corollary 5.3:** *Algorithm 3.1 coupled with the updating scheme 5.1 solves problems (P) and (D) in no more than $O(n^3 L)$ arithmetic operations.*

**Proof:** From corollary 4.4, we know that algorithm 3.1 finds an optimal solution in $O(\sqrt{n}L)$ iterations. Proposition 5.2 implies that the total number of rank-one updates is then of the order of $O(nL)$. Since each rank-one update involves $O(n^2)$ arithmetic operations, the total number of arithmetic operations is then of the order of $O(n^3L)$. This completes the proof of the corollary. $\square$

### 6. Initialization of the Algorithm

Given an LP in standard form and its dual, we have assumed in section 2 that conditions (a) and (b) of assumption 2.1 were valid. In general, this is not the case. In this section, we show how to transform an LP problem in standard form

$$(\tilde{P}) \quad \min \ \tilde{c}^T v$$

$$s.t. \ \tilde{A}v = \tilde{b} \ , \qquad \tilde{A} \ \textit{full row rank}$$

$$v \geq 0$$

to an equivalent LP problem in standard form so that the new problem and its dual satisfy the assumption 2.1. As a result of the process to be described below, we will also obtain an initial point satisfying the criterion of closeness (3.2). The algorithm can then start from this initial point.

Let m and n be integers such that $\tilde{A} \in \mathbf{R}^{(m-1)\times(n-2)}$. First we scale the problem $(\tilde{P})$ by introducing a change of variables $v = \lambda\tilde{x}$ where $\lambda > 0$ is a scaling factor and $\tilde{x} \in \mathbf{R}^{(n-2)}$. The scaling factor $\lambda$ is chosen large enough so that any basic feasible solution $v$ of $(\tilde{P})$ satisfies $e^T v \leq (n-2)\lambda$. We then obtain the following equivalent problem

$$\min \ (\lambda \tilde{c}^T)\, \tilde{x}$$

$$s.t. \ (\lambda \tilde{A})\, \tilde{x} = \tilde{b}$$

$$e^T \tilde{x} + \tilde{x}_{n-1} = n-1$$

$$\tilde{x} \geq 0 \ , \ \tilde{x}_{n-1} \geq 0$$

where $\tilde{x}_{n-1}$ is a slack variable.

Now we want the vector of ones to be a feasible solution of the new problem. We introduce an artificial variable $\tilde{x}_n$ with a large cost M to obtain

$$\min \ (\lambda \tilde{c}^T)\, \tilde{x} + M\, \tilde{x}_n$$

$$s.t. \ (\lambda \tilde{A})\, \tilde{x} + (\tilde{b} - \lambda \tilde{A} e)\, \tilde{x}_n = \tilde{b}$$

$$e^T \tilde{x} + \tilde{x}_{n-1} + \tilde{x}_n = n$$

$$\tilde{x} \geq 0 \ , \ \tilde{x}_{n-1} \geq 0 \ , \ \tilde{x}_n \geq 0$$

Letting $B = [\lambda \tilde{A} \mid \tilde{b} - \lambda \tilde{A} e \mid 0] \in \mathbb{R}^{(m-1) \times n}$, $c^T = (\lambda \tilde{c}^T, 0, M)$ and $x = (\tilde{x}, \tilde{x}_{n-1}, \tilde{x}_n) \in \mathbb{R}^n$, we can rewrite the previous problem as

$$(P) \quad \min \ c^T x$$

$$s.t. \ Bx = \tilde{b}$$

$$e^T x = n$$

$$x \geq 0$$

We can recast this problem in the notation of problem (P) of section 2 just by letting $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be

$$A = \begin{bmatrix} B \\ e^T \end{bmatrix} \qquad b = \begin{pmatrix} \tilde{b} \\ n \end{pmatrix}$$

We now show that this problem satisfies assumption 2.1 of section 2. Note that, by the construction above, the vector of ones $e \in \mathbf{R}^n$ is feasible for problem (P). This shows that problem (P) satisfies (a) of assumption 2.1. By assumption, the matrix $\tilde{A}$ has full row rank. This immediately implies that the matrix $B$ has full row rank. Since $Be = 0$, it follows that the matrix $A$ has full row rank. Therefore, problem (P) satisfies (c) of assumption 2.1. To see that problem (P) also satisfies condition (b) of assumption 2.1, consider its dual

$$(D) \quad \max \ \tilde{b}^T \eta + nt$$

$$s.t. \ B^T \eta + et + z = c$$

$$z \geq 0$$

where $\eta \in \mathbf{R}^{(m-1)}$, $t \in \mathbf{R}$ and the slack $z \in \mathbf{R}^n$. Problem (D) has a feasible point with the corresponding slack $z$ strictly positive. To see this, set $\eta = 0$ and choose any $t$ satisfying

$$t < \min_i c_i$$

Hence condition (b) of assumption 2.1 is satisfied. We will now obtain an initial point satisfying the criterion of closeness (3.2). Let $x^o = e$, $\eta^o = 0$, and $t^o = -\mu^o$ where $\mu^o$ satisfies

$$\mu^o \geq \frac{\|c\|}{\theta} \tag{6.1}$$

Then $x_j^o z_j^o = c_j + \mu^o$, $j=1, \ldots, n$ and the criterion for closeness (3.2) becomes

$$\left( \sum_{j=1}^{n} c_j^2 \right)^{1/2} \leq \theta \, \mu^o$$

which is satisfied due to (6.1). In summary, we can apply the algorithm of section 3 to solve the pair of problems (P) and (D) with the initial point $w^o \equiv (x^o, y^o, z^o)$ determined by

$$x^o = e$$

$$y^o \equiv (w^o, t^o) = (0, -\mu^o)$$

$$z^o = c + \mu^o e$$

where $\mu^o$ satisfies (6.1).

## 7. Remarks and Extensions

The following observations are in order:

(1) The purpose of this paper is to present a theoretical result. Thus in order to simplify the presentation, we constructed $\hat{\mu} = \mu(1 - \delta/\sqrt{n})$. Obviously, one can use $\hat{\mu}$ which is less than or equal than the above one, but still satisfying (b) of theorem 4.1 and relations (4.7) and (4.8). In this way, one can accelerate the convergence of the algorithm.

(2) Additional improvements in actual implementation, which are possible, such as more judicious selection of $\theta$, $\delta$ and $\gamma$, together with actual test results, are the subject of a forthcoming paper.

### 7.1. Extensions to Convex Quadratic Programming

The same ideas presented here can be applied to the convex quadratic programming as follows. Let

$$(Q) \quad \min \; c^T x + \frac{1}{2} x^T Q x$$

$$s.t. \; Ax = b$$

$$x \geq 0$$

where $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite. Using the same approach and following [6] one can define the following barrier problem

$$(Q\mu) \quad \min \; c^T x + \frac{1}{2} x^T Q x - \mu \sum_{j=1}^{n} \ln x_j$$

$$s.t. \; Ax = b$$

$$x > 0$$

with the Kuhn-Tucker conditions:

$$(i) \; ZXe - \mu e = 0$$

$$(ii) \; Ax - b = 0 \, , \, x > 0$$

$$(iii) \; A^T y + z - c - Qx = 0$$

As in the linear case the path of solutions to $(Q\mu)$ converges to the optimal solution of $(Q)$ as $\mu$ tends to zero (see Meggido [6]). Similarly to algorithm 3.1, given a point $w = (x, y, z)$, one can construct $\Delta w = (\Delta x, \Delta y, \Delta z)$ by solving the following system of linear equations

$$Z\Delta x + X\Delta z = XZe - \hat{\mu} e$$

$$A\Delta x = 0$$

$$A^T \Delta y + \Delta z - Q\Delta x = 0$$

and then proceed to $\hat{w}$ by

$$\hat{w} = w - \Delta w$$

It can be shown that convergence properties of this algorithm are the same as the ones of algorithm 3.1 ( i.e., $O(\sqrt{n}L)$ iterations, each of which requires at most $O(n^3)$ arithmetic operations). The proofs are slight modifications of the proofs of this paper. The details of this algorithm together with implementation tests are the subject of a forthcoming paper.

## 8. Acknowledgement

This research was partially funded by the Brazilian Post-Graduate Education Federal Agency (CAPES) and by the United States Navy Office of Naval Research under contract N00014-87-K-0202. Their financial support is gratefully acknowledged.

## References


[1] K.R. Frisch, "The Logarithmic Potential Method of Convex Programming", *unpublished manuscript, University Institute of Economics,* Oslo, Norway, 1955

[2] P.E. Gill, W. Murray, M.A. Saunders, J.A. Tomlin, M.H. Wright, "On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar's Projective Method", *Technical Report Sol 85-11, Systems Optimization Laboratory, Department of Operations Research, Stanford University,* Stanford 1985

[3] C.C. Gonzaga, "An Algorithm for Solving Linear Programming Problems in $O(n^3L)$ Operations", *Memorandum Number UCB/ERL M87/10, Electronics Research Laboratory, University of California,* Berkeley, CA March, 1987

[4] N. Karmarkar, "A New Polynomial Time Algorithm for Linear Programming", *Combinatorica,* vol.4, 373-395 1984

[5] M. Kojima, S. Mizuno, A. Yoshise, "A Primal-Dual Interior Point Algorithm for Linear Programming", *Report No. B-188, Department of Information Sciences, Tokyo Institute of Technology,* Tokyo, Japan February, 1987

[6] N. Meggido, "Pathways to the Optimal Set in Linear Programming", *Research Report, IBM Almaden Research Center,* San Jose, CA 1986

[7] J. Renegar, "A Polynomial-Time Algorithm Based on Newton's Method For Linear Programming", *Report MSRI 07 118-86, Mathematical Sciences Research Institute,* Berkeley,California, June 1986

[8] P.M. Vaidya, "An Algorithm For Linear Programming Which Requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ Arithmetic Operations", *preprint, AT&T Bell Laboratories,* Murray Hill, N.J. 1987

END
9-87
DTIC